一种基于模式匹配度的用户移动规则挖掘及位置预测方法研究 *

张海涛 [a]，蒋继飞 [b]，周 欢 [b]

(南京邮电大学 a. 地理与生物信息学院; b. 通信与信息工程学院，南京 210003)

摘 要：移动用户的位置预测可以有效地用于系统资源的时空分配，提高移动通信系统的资源利用率。传统的位置预测方法由于模式支持度计算方式不合理，存在预测精度偏低的问题。为此，提出了一种基于模式匹配度的用户移动规则挖掘及位置预测方法，并将其用于移动通信系统中，以基台覆盖范围网格为单元的用户位置预测。具体包括三个步骤：通过图的遍历挖掘用户移动模式、基于用户移动模式生成用户移动规则和依据用户移动规则进行位置预测。实验分析使用 10 个批次轨迹数据进行用户移动规则挖掘，结果表明，该方法挖掘出的用户移动规则数少、支持度高和置信度高，具有高精度的优点。

关键词：移动用户；位置预测；模式匹配度；用户移动规则

中图分类号：TP311　　doi: 10.3969/j.issn.1001-3695.2018.04.0280

Method of mining user mobile rule based on pattern matching degree and location prediction.

Zhang Haitao[a], Jiang Jifei[b], Zhou Huan[b]

(*a. College of Geographic & Biologic Information, b. College of Communication & Information Technology, Nanjing University of Posts & Telecommunications, Nanjing* 210003, *China*)

Abstract: The location prediction of mobile user is effective for the temporal and spatial allocation of system resources, and it can improve the resource utilization of the mobile communication system. The traditional location prediction method has a problem of low prediction accuracy due to unreasonable calculation of support. To overcome these challenges, this paper presented a method called mining user mobile rule based on pattern matching degree and location prediction. It can predict location with the base coverage mesh as the unit for mobile users in mobile communication systems. The detailed procedure contained three phases: a) mining user mobile patterns from graph traversal; b) generating user mobile rules from user mobile patterns; c) predicting based on user mobile rules. Finally, the conducted experiments on 10 batches of trajectory data demonstrate that this method has the characteristics of fewer rules, high support, and high confidence than traditional methods, and highly predict accuracy.

Key words: mobile user; location prediction; pattern matching degree; user mobile rule

0 引言

随着计算机技术和通信技术的发展，移动通信系统日益普及[1,2]。移动计算为人们在任何时间地点访问任何数据的需求提供了实现手段。它被视为计算机的热点领域[3,4]。移动计算具有移动性、网络多样性等特点[5~7]。目前，以 3G、4G 为主的移动通信系统，具有服务用户庞大及数据服务类型多样性（如视频、声音、图像等）的特点。为保障服务的质量，需要对移动通信系统中的移动用户信息进行更加有效的管理。移动性位置管理也即移动用户的位置信息存储、更新和预测是移动通信系统管理的主要内容[8~11]。其中，移动用户的位置预测可以有效地用于系统资源的时空分配，以提高移动通信系统的资源利用率，并降低移动用户在基台覆盖范围网格（mesh）之间切换（handover）时产生的系统延迟及掉线[12~14]，最终能够明显的提高移动通信网络的服务质量。

针对位置预测国内外学者提出了很多解决方法。文献[15,16]提出了一种 GMM 和 LMM 相结合的方法，使用 GMM 与 LMM 分别建立用户在基台覆盖范围网格间和基台覆盖范围网格内的移动模型，并基于移动用户的运动轨迹与移动模型的匹配，对移动用户的位置进行预测。文献[17]Jeung Hoyoung 和 Qing Liu 等人利用 Apriori 算法对用户移动轨迹进行处理，提取出用户移动模式以此来进行用户位置预测。文献[18,19]中，

收稿日期：2018-04-09；**修回日期**：2018-05-29　**基金项目**：国家自然科学基金资助项目（41201465）；江苏省自然科学基金资助项目（BK2012439）；江苏省社会发展项目（BE2016774）

作者简介：张海涛（1978-），男，副教授，博士，主要研究方向为移动 GIS 理论方法与关键技术、时空数据挖掘、LBS 隐私保护等（zhanghaitao@njupt.edu.cn）；蒋继飞（1993-），男，硕士，主要研究方向为数据挖掘、隐私保护；周欢（1993-），男，硕士，主要研究方向为隐私保护.

Morzy 利用 Aprior 算法进行改进后生成关联规则，后来又依据改进的 PrefixSpan 算法提取出用户移动的频繁模式，通过对用户的频繁模式进行处理生成预测位置的用户规则。

但是上述方法存在共性问题是在用于用户位置预测的模式或规则其支持度计算方法都是采用简单的计数方式，用户位置预测的精度较低。不能对用户位置预测的高精度问题进行有效处理。而高精度是保证移动通信系统为移动用户高效分配系统资源的关键技术问题。因此本文提出一种基于模式匹配度的用户移动规则挖掘及位置预测方法研究。

1　位置预测方法的基本原理

首先，给出相关方法涉及的定义。

定义 1　$G=<V,E>$ 表示移动通信系统中所有基台覆盖范围形成的有方向性的，无权重的图结构。假设 $T=<C_1,C_2,C_3...C_n>$ 表示用户在基台覆盖范围内的移动路径即移动序列，每个 C_K 网格，且连续的网格之间空间邻近。序列数据库 D 为 $<sid,T>$ 的元组集合，其中 sid 为当前移动序列的 id，T 表示移动序列。假设 L_K 为根据序列数据库挖掘出长度为 K 的用户移动模式集合，对每一个长度为 K 的用户移动模式检查其最后一个网格是否在 G 中有空间临近点，若有，将其添加到该用户移动模式，最终形成长度为 K+1 的候选模式集合 C_{K+1}。根据定义 3 定义的支持度计算方式，计算出序列数据库 D 对候选模式集合 C_{K+1} 的支持度。设定支持度阈值为 $Sup_{\min}$，去掉候选模式集合 C_{K+1} 中支持度小于 $Sup_{\min}$ 候选模式，生成长度为 K+1 的用户移动模式集合 L_{K+1}。由此逐层搜索得出所有用户移动模式集合 L。

定义 2　如果 x 和 y 都是单独的字符或者空格，$\delta(x,y)$ 表示匹配 x 和 y 后的得分函数，公式定义为

$$\delta(x,y)=\begin{cases}0, & x=y\\ 1, & 其他\end{cases} \tag{1}$$

如果 S 是一字符串，$|S|$ 表示 S 的长度，$S[i]$ 代表字符串 S 中的第 i 个字符（第一个字符是 S[1]，而不是 S[0]）。假设 A 是一条移动用户在运动时产生的以基台覆盖范围网格（mesh）为单元的路径（user path based mesh,UPM）（简称用户移动路径），UPMB 就是包含在 A 里的一个用户移动模式（user mobile pattern,UMP）。对齐匹配排列 X' 把 A 和 B 写成可能包含空格的字符串 A' 和 B'。其中：$|A'|=|B'|$；分别从 A' 和 B' 中去除空格后就剩下 A 和 B。那么满足该特征的一个对齐匹配排列 X'，其对齐匹配得分函数的公式定义为 $\delta(X')=\sum_{i=k}^{m}\delta(A'[i],B'[i])$。其中 k、m 分别是 B' 中第一个和最后一个非空格字符的索引号。排列 X' 的对齐匹配得分值实际上就是 B' 中第一个和最后一个非空格字符之间错误匹配字符的个数。因此，最佳的对齐匹配排列 X' 应具有最小的对齐匹配得分值即最小的模式匹配度，我们将该最小得分称为 totDist。用户移动路径（UPM）A 对用户移动模式（UMP）的增量支持度（suppInc）的定义公式为

$$suppInc(A,B)=\begin{cases}\dfrac{1}{1+totDist}, & \text{UPM A包含UMP B}\\ 0, & 其他\end{cases} \tag{2}$$

定义 3　D 是存储所有用户移动路径即移动序列的数据库，B 是数据库 D 中包含的一个用户移动模式（UMP），则数据库 D 对 B 的支持度定义为

$$supp(B)=\sum suppInc(\alpha,B),\alpha\in D \tag{3}$$

定义 4　定义 $M=<m_1,m_2,...,m_{k-1},m_k>$，k>1 表示长度为 k 的用户移动模式（UMR），则定义用户移动规则（UMR）为 $R_i=<m_1,m_2,...,m_i>\to<m_{i+1},m_{i+2},...,m_k>(1\le i<k)$，那么根据用户移动模式推导出的用户移动规则有：

当 i=1 时，$R_1=<m_1>\to<m_2,m_3,...,m_k>$；

当 i=2 时，$R_2=<m_1,m_2>\to<m_3,m_4,...,m_k>$；

...

当 i=k-1 时，$R_{k-1}=<m_1,m_2,...,m_{k-1}>\to<m_k>$；

其中:箭头 $(\to)$ 前面为规则头部，后面为规则尾部，定义该用户移动规则的置信度为

$$conf(M)=\frac{supp(<m_1,m_2,...,m_k>)}{supp(<m_1,m_2,...,m_{i-1}>)}\times 100\% \tag{4}$$

其中:supp(M)表示为用户移动模式支持度。设定置信度阈值为 $conf_{\min}$。在产生所有可能的用户移动规则和其对应的置信度时，对于 $conf(M)\ge conf_{\min}$ 的用户移动规则保存用来预测用户位置。

定义 5　给定用户移动轨迹 $P=<m_1,m_2,...,m_{i-1}>$，m_i 为用户所在的网格。用户移动规则为 $R_j=<a_1,a_2,...,a_j>\to<a_{j+1},...,a_k>$ 用户移动规则集合为 R 则 $R_j\subset R$。若 R_j 的头部包含于 P 中，且头部最后一个元素 a_j 与 P 的最后一个元素 m_{i-1} 重合。则定义该用户移动规则为匹配规则，并且计算该匹配规则的支持度与置信度的和。那么据此定义可以从用户移动规则集中筛选出匹配规则集，并且依据匹配规则的支持度与置信度和对匹配规则集进行降序。假定 $R_j=<a_1,a_2,...,a_j>\to<a_{j+1},...,a_k>$ 为经过降序处理的匹配规则集中首个匹配规则，则 a_{j+1} 即为用户移动轨迹 P 的下一个预测位置网格。

2　基于用户移动规则的预测方法

位置预测包含三个阶段：基于图的遍历挖掘出用户移动模式、从用户移动模式生成用户移动规则、基于用户移动规则进行位置预测。

2.1　从图的遍历中挖掘用户移动模式

该阶段包括三个步骤，具体过程如下：

a）将研究的移动通信系统中的所有基台覆盖范围网格（Mesh），转换成以有方向性的、无权重的图（graph）结构 $G=<V,E>$ 存储表达的数据;b）收集以基台覆盖范围网格

（mesh）为单元的所有用户移动路径（UPM）数据，检查每条用户移动路径（UPM）中的连续网格元素（mesh）是否空间临近，并在序列数据库 D 中存储;c）设定挖掘用户移动模式（UMP）的支持度阈值 $(supp_{\min})$，使用算法 $UMPMining(G,D,supp_{\min})$ 从序列数据库 D 中挖掘出长度逐级增加的用户移动模式（UMP）集合。其中，使用增量支持度对候选模式进行支持度计数、使用支持度阈值从候选模式获取用户移动模式（UMP）、使用空间临近性搜索从长度为 k 的用户移动模式（UMP）得到长度为(k+1)的候选模式（算法实现 CandidateGeneration()）

算法 1　UMPMining()

输入：存储所有用户移动路径（UPM）的数据库 D、有基台覆盖范围网格（mesh）对应的图结构 ,V 是 G 中所有节点集合，E 是连接 G 中所有节点 V 的有向边的集合、挖掘用户移动模式（UMP）的支持度阈值 $supp_{\min}$。

输出：所有的用户移动模式（UMP）集合 L。

1．$C_1 \leftarrow$ 长度为 1 的候选模式

2． k=1

3．$L = \varnothing$ //用户移动模式（UMP）集合，初始值设置为空

4．$While\ C_K \neq \varnothing$ {

5．$ForeachUPM\ \ \alpha \in D$ {

//c 是长度为 k 的候选模式，同时也是用户移动路径（UPM）α 的子序列

6．$C = \{s \mid s \in C_K \ \&\&s \subseteq \alpha\}$

7．$Foreach$ c∈C{c.count=c.count+s.suppInc} //使 c 的支持度递增

8． }//得到长度为 k 的用户移动模式（UMP）

9．$L_K = \{c \mid c \in C_K \ \&\&c.count \geq supp_{\min}\}$

10．$L = L \cup L_k$ //把 L_k 添加到 L 中

//生成长度为 k+1 的候选模式

11．$c_{k+1} \leftarrow CandidateGeneration(L_k, G), \forall c \in C_{k+1}$

$c.count = 0$

12．k = k +1}

13．return L

其中，算法 $UMPMining(G,D,supp_{\min})$ 中第 5~8 行利用增量支持度，对候选模式进行支持度计数（对应定义 2、定义 3）；第 9、10 行利用候选模式的支持度计数（c.count）与设定的支持度阈值 $(supp_{\min})$ 的比较，从候选模式集合中得到用户移动模式（UMP）集合；第 11 行内嵌的算法 $CandidateGeneration(L_k,G)$，该算法通过使用空间临近性搜索从长度为 k 的用户移动模式（UMP）得到长度为(k+1)的候选模式。

算法 2　CandidateGeneration ()

输入:长度为 k 的用户移动模式（UMP）L_k、所有基台覆盖范围网格（Mesh）对应的图结构 $G=<V,E>$,V 是 G 中所有节点集合，E 是连接 G 中所有节点 V 的有向边的集合。

输出：长度为 k+1 的候选模式集合 candidates。

1. $Candidates = \varnothing$ //候选模式集合，初始值设置为空

2. $Foreach\ L =< l_1, l_2, \ldots, l_k >, L \in L_k$ {

/*对每个长度为 k 用户移动模式（UMP）L，获取其成员 l_k 在 G 中的邻近覆盖区域网格*/

3. $N^+(l_k) = \{v \mid v \in V, e \in E, \exists e(l_k, v)\}$

/*在 G 中查找所有的节点 v，其同 l_k 在 G 中的节点具有有向连接边 e。因此，$N^+(l_k)$ 是用户从 l_k 出发可能到达邻近覆盖区域网格（Mesh）的集合*/

4. $Foreach\ v \in N^+(l_k)$ {

/*对满足条件的所有节点 v，将其添加为模式 L 最后一个元素，从而生成相应的候选模式*/

5. $C^{'} =< l_1, l_2, \ldots, l_k, V >$

//将 $C^{'}$ 添加到候选模式集合 candidates 中

6. $Candidates \leftarrow Candidates \cup C^{'}$ }}

7. return candidates

2.2 从用户移动模式生成用户移动规则

该阶段包括四个步骤，具体过程如下：

a）设定挖掘用户移动规则（UMR）的置信度阈值 $(conf_{\min})$;b）选择长度大于 1 的所有用户移动模式（UMP），生成一系列的用户移动规则（UMR）;c）扫描序列数据库 D 中的用户移动路径（UPM）数据，对生成的系列用户移动规则（UMR），进行置信度的统计计算;d）按照设定的置信度阈值 $(conf_{\min})$，逐一比较生成用户移动规则（UMR）的置信度，从而得到满足条件的用户移动规则集合 R。

2.3 基于用户移动规则的位置预测

该阶段包括三个步骤，具体过程如下：

a）取用于预测移动用户位置的、当前的用户移动路径（UPM）P;b）设定预测移动用户下一步可能到达的基台覆盖范围网格（mesh）的数量（m）;c）利用算法 $MobilityPrediction(P,R,m)$ 对移动用户的下一步位置进行预测。

算法 3　$MobilityPrediction()$

输入：用户当前的移动轨迹 $P =< m_1, m_2, \ldots, m_{i-1} >$、用户移动规则（UMR）集合 R、预测用户下一步可能到基台覆盖范围网格（mesh）的数量 m。

输出：预测到达的基台覆盖范围网格（Mesh）集合，PMeshs。

1．$PMeshs = \varnothing$ //初始结果设置为空

2． k = 1

//检查 R 中的所有规则

3．$ForeachUMR\ r{:}< a_1, a_2, \ldots, a_j > \rightarrow < a_{j+1}, \ldots, a_k > \in R$　{

/*找出匹配规则的集合：如果规则的头部（head）包含于 P 中，且规则的头部（head）的最后一个元素 (a_j) 与 P 的最后一个元素 (m_{i-1})（即用户当前所在的网格）重合，则该规则为匹配规则*/

4．$If < a_1, a_2, \ldots, a_j > \subset P \ \&\&a_j = m_{i-1}$ {

/*把匹配规则的尾部（tail）的第一个元素，以及匹配规则的置信度与支持度的和（同时考虑度量规则有效性的两个指标：supp，conf）保存到数组变量中*/

5. $TupleArray[k]=(a_{j+1}, r.conf+r.supp)$

6. $k=k+1$ }}

//按照置信度与支持度的和，对数组变量进行降序

7. $TupleArray \leftarrow sort(TupleArray)$

8. $index=0$

9. $while(index<m \& \& index<TupleArray.length)\{$

//从数组变量中选择前 m 个元素

10. $PMeshs \leftarrow PMeshs \cup TupleArray[index]$

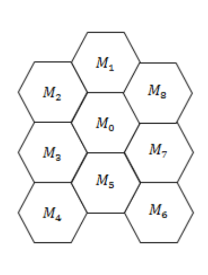
11. $index=index+1$ }

12. Return PMeshs

其中，算法 $MobilityPrediction(P,R,m)$ 中第 3~6 行利用用户移动路径（UPM）P 与用户移动规则（UMR）的匹配运算获取系列的匹配规则；第 8 行按照匹配规则的支持度、置信度，这两个衡量规则有效型的指标进行规则排序；第 9~11 行从排序后的匹配规则中选择前 m 个结果；第 12 行返回最终的规则结果集合 PMeshs ，其中包含的基台覆盖范围网格（Mesh）即为预测的结果。

2.4 实例分析

下面结合一个例子，介绍算法执行的基本过程。将图 1 所示的 9 个基台覆盖范围网格转换成对应的图结构，并检查表 1 所示的所有的用户移动路径（UPM）数据，并确定每条用户移动路径（UPM）中的连续网格元素均是空间临近，并将表 1 的结果保存到数据库 D 中。



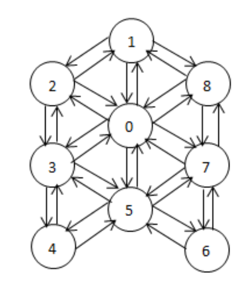


图 1 所有基台覆盖范围网格及其对应的图结构

表 1 存储用户移动路径的数据库

ID	UPM
1	<7,6,5,0>
2	<2,3,0,1>
3	<1,2,3,4,5,7,0>
4	<5,0,7,0,1>

设定挖掘用户移动模式（UMP）的支持度阈值 $(supp_{min}=2)$，并使用算法 $UMPMining()$ 从序列数据库 D 中挖掘出长度逐级增加的用户移动模式（UMP）集合。

结果如表 2~5 所示，C 表示候选模式集合，L 表示用户移动模式集合，数字表示长度。

表 2 长度为 1 的候选模式集合和长度为 1 的用户移动模式集合

C1		L1	
CAND	SUPP	PATTERN	SUPP
<0>	4	<0>	4
<1>	3	<1>	3
<2>	2	<2>	2
<3>	2	<3>	2
<4>	1	<5>	5
<5>	5	<7>	3
<6>	1		
<7>	3		
<8>	0		

表 3 长度为 2 的候选模式集合和长度为 2 的用户移动模式集合

C2						L2	
CAND	SUPP	CAND	SUPP	CAND	SUPP	PATTERN	SUPP
<0,1>	2	<2,1>	0.33	<5,4>	0	<0,1>	2
<0,2>	0	<2,0>	0.7	<5,3>	0	<0,7>	2
<0,3>	0	<2,3>	2	<5,0>	2.5	<2,3>	2
<0,5>	0	<3,2>	0	<5,7>	2	<5,0>	2.5
<0,7>	2	<3,0>	1.25	<5,6>	0	<5,7>	2
<0,8>	0	<3,5>	0.5	<7,6>	1	<7,0>	2.33
<1,2>	1	<3,4>	1	<7,5>	0.5		
<1,0>	0.17			<7,0>	2.33		
<1,8>	0			<7,8>	0		

表 4 长度为 3 的候选模式集合和长度为 3 的用户移动模式集合

C3						L3	
CAND	SUPP	CAND	SUPP	CAND	SUPP	PATTERN	SUPP
<0,1,2>	0	<5,0,1>	0	<7,0,1>	1	<5,0,7>	2
<0,1,0>	0	<5,0,2>	0	<7,0,2>	0		
<0,1,8>	0	<5,0,3>	0	<7,0,3>	0		
<0,7,8>	0	<5,0,5>	0	<7,0,5>	0		
<0,7,0>	0	<5,0,7>	2	<7,0,8>	0		
<0,7,5>	0	<5,0,8>	0				
<0,7,6>	0	<5,7,8>	0				
<2,3,2>	0	<5,7,0>	1				
<2,3,0>	1.25	<5,7,5>	0				
<2,3,5>	0.5	<5,7,6>	0				
<2,3,4>	1						

表 5 长度为 4 的候选模式集合

C4	
CAND	SUPP
<5,0,7,8>	0
<5,0,7,0>	0
<5,0,7,5>	0
<5,0,7,6>	0

根据表 2~5 的结果最终得到长度分别为 1，2，3 的用户移

动模式（UMP）集合及对应的支持度，如表 6 所示。

表 6　所有用户移动模式(UMP)集合

L			
CAND	SUPP	CAND	SUPP
<0>	4	<0,1>	2
<1>	3	<0,7>	2
<2>	2	<2,3>	2
<3>	2	<5,0>	2.5
<5>	5	<5,7>	2
<7>	3	<7,0>	2.33
		<5,0,7>	2

设定挖掘用户移动规则（UMR）的置信度阈值 $(conf_{min}=50)$。从表 6 中选择长度大于 1 的用户移动模式（UMP），生成一系列的用户移动规则（UMR）。扫描表 1 中的用户移动路径（UPM）数据，按照定义 4 对生成的系列用户移动规则（UMR），进行置信度的统计计算。

所有可能的用户移动规则（UMR）及置信度的结果如表 7 所示。按照设定的置信度阈值 $(conf_{min}=50)$，得到最终的用户移动规则（UMR）如表 8 所示。

表 7　所有可能的用户移动规则(UMR)

Rule	Conf
<0> → <1>	50
<0> → <7>	50
<2> → <3>	100
<5> → <0>	5
<5> → <7>	40
<7> → <0>	77.67
<5,0> → <7>	80
<5> → <0,7>	40

表 8　满足置信度阈值的用户移动规则(UMR)

Rule	Conf
<0> → <1>	50
<0> → <7>	50
<2> → <3>	100
<5> → <0>	50
<7> → <0>	77.67
<5,0> → <7>	80

获取用于预测移动用户位置的、当前的用户移动路径 $(UPM)P=<3,5,7,0>$。设定预测移动用户下一步可能到达的基台覆盖范围网格（Mesh）的数量（m=1）。利用算法 $MobilityPrediction(P,R,m)$ 对移动用户的位置进行预测。根据表 9 得出预测结果为 7 的覆盖区域网格。

表9 用于预测的用户移动规则（UMR）

Rule	Conf+ SUPP
<5,0> → <7>	82
<0> → <1>	52
<0> → <7>	52

3　实验结果与分析

3.1　实验数据

本文初始数据为电信数据中的请求位置数据，其通过 60 s 采集一次，时间范围为 24 h。一个用户的请求位置数据代表了该用户移动过程中经过的基台覆盖范围网格，通过空间离散化将用户移动的网格序列变为空间临近的连续网格序列。本文使用了 10 个批次轨迹数据基本信息如表 10 所示。

表 10　10 个批次轨迹数据基本信息

不同批次	用户数量	轨迹点总数量	轨迹点总数量去重	移动轨迹平均轨迹点数
1	200	231779	17838	1158
2	201	240375	17455	1195
3	200	236833	19522	1184
4	198	231857	19236	1170
5	202	237104	19021	1175
6	201	231732	20130	1153
7	200	237718	20460	1188
8	199	239059	21653	1201
9	203	236350	22474	1164
10	200	233262	19447	1166

3.2　性能对比实验

通过前面对传统预测方法研究发现，本文提出方法与传统方法存在的本质区别点在于支持度计算方式不一样，为了避免与传统方法逐一对比，同时使对比更具有公平性，我们采用本文设计方法的基本框架，设计与传统的方法一样支持度计算方式的挖掘预测方法。传统计算支持度方式，即采用 0 或 1 的度量方法：

$$suppInc(A,B)=\begin{cases}1, & UPM\ A\ 包含\ UMP\ B\\ 0, & 其他\end{cases}$$

实验设定挖掘算法的支持度阈值，置信度阈值分别为 10%，50%。

在后面的实验及结果分析中，我们简称基于模式匹配度计算支持度的方法为相似化方法，而采用 0 或 1 计算支持度的对比方法为非相似化方法。相似化方法和非相似化方法挖掘用户移动规则的基本性能指标如图 2~图 4 所示，从这 10 个批次可以发现本文设计的相似化方法挖掘的用户移动规则数量少，支持度高，置信度高的特点。数量少是因为本文方法重新定义了用户移动路径对用户移动模式支持度使得挖掘出的用户移动模式集合变少导致挖掘用户移动规则变少。支持度高是因为相似化方法挖掘出的用户移动规则数量少。置信度均值高是因为对比方法（非相似化）采用简单的 0 或 1 的度量方法计算支持度，导致产生了大量的虚假的强规则。

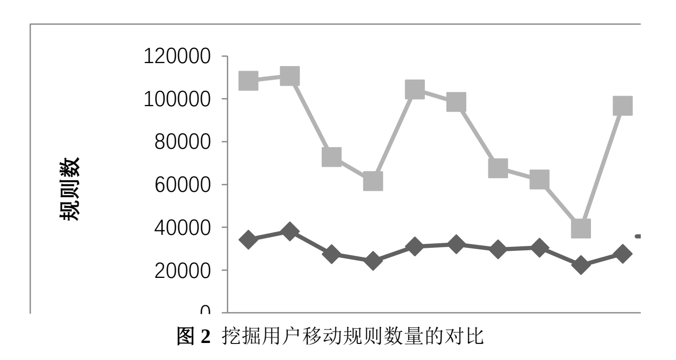


图 2 挖掘用户移动规则数量的对比

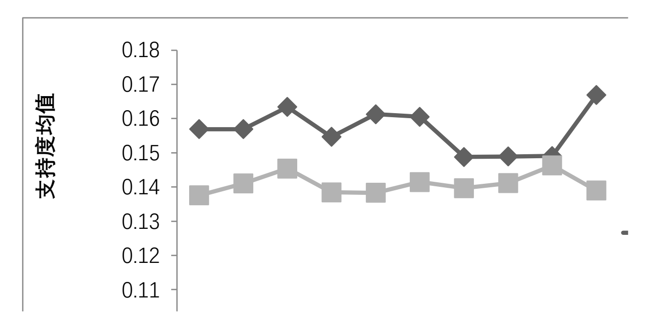


图 3 挖掘用户移动规则支持度均值对比

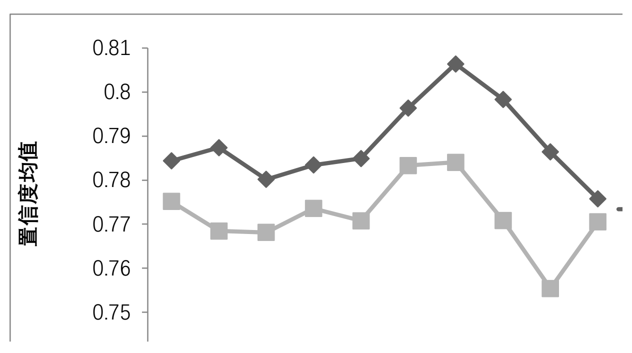


图 4 挖掘用户移动规则置信度均值对比

相似化与非相似化方法挖掘用户移动规则预测指标对比如图 5~7 所示。可以看出，相似化方法挖掘出的用户移动规则具有较高的准确率均值、召回率均值和 F 均值。准确率均值高说明通过相似化方法挖掘出的用户移动规则在用于测试数据预测时具有更高的预测精度。召回率均值高说明通过相似化方法挖掘出的用户移动规则在用于测试数据预测时具有更多匹配到的规则数量。F 值均值是准确率和召回率加权（权值 a=1）调和平均，这能够反映出挖掘用户移动规则的预测性能。F 值高说明采用本文设计的相似化方法挖掘用户移动规则能够有比较高的可用性。

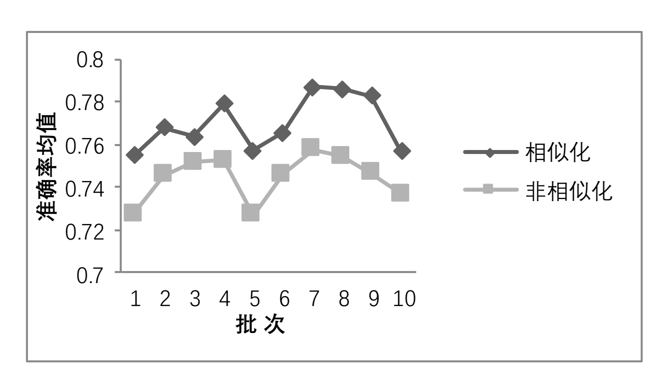


图 5 挖掘用户移动规则准确率均值对比

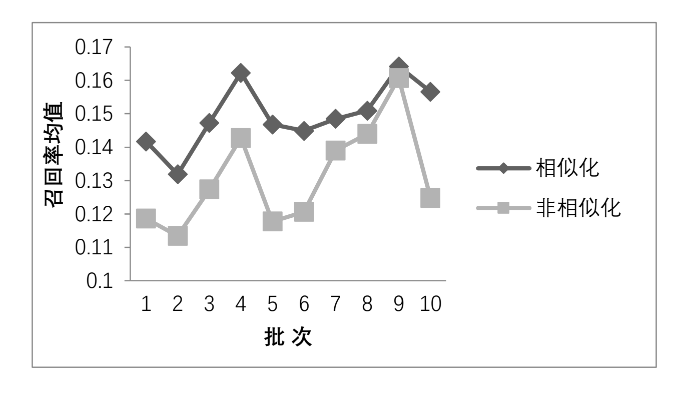


图 6 挖掘用户移动规则召回率均值对比

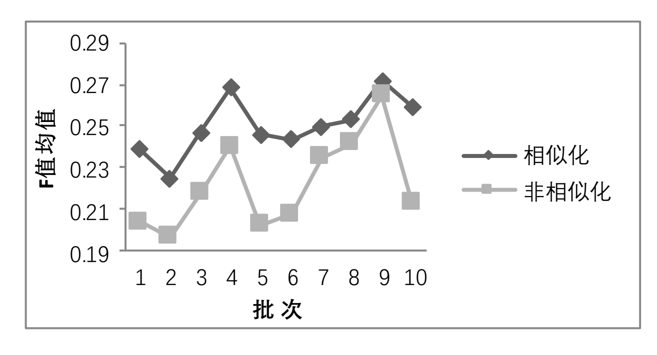


图 7 挖掘用户移动规则 F 值均值对比

4 结束语

移动用户的位置预测，可以有效地用于系统资源的时空分配，成为移动通信系统中资源管理的重要内容之一。预测准确度是位置预测性能的重要指标。但是，传统的位置预测方法并不能有效应对此类问题。本文提出一种基于模式匹配度的用户移动规则挖掘及位置预测方法研究。该方法通过定义度量用户移动路径（UPM）与用户移动模式（UMP）匹配度得分值，获取用户移动路径（UPM）与用户移动模式（UMP）的最佳匹配方法，从而设计了一种更加合理的度量用户移动路径（UPM）对用户移动模式（UMP）支持程度的计算方法。通过实验结果表明：本文提出的方法挖掘出的用户移动规则具有规则数少、支持度高和置信度高的特点，在用于预测数据时能够具有更高的预测精度。

参考文献：

[1] Dinh H T, Lee C, Niyato D, *et al.* A survey of mobile cloud computing: architecture, applications, and approaches [J]. Wireless Communications & Mobile Computing, 2013, 13 (18): 1587-1611.

[2] Park Y I, Jang S M. Understanding privacy knowledge and skill in mobile communication [J]. Computers in Human Behavior, 2014, 38 (9): 296-303.

[3] 胡海洋；刘润华；胡华. 移动云计算环境下任务调度的多目标优化方法 [J]. 计算机研究与发展，2017，54 (9)：1909-1919. (Hu Haiyang, Liu Runhua, Hu Hua. Multi-objective optimization for task scheduling in mobile cloud computing [J]. Journal of Computer Research and Development, 2017, 54 (9): 1909-1919.)

[4] 郎为民. 移动云计算：架构、算法与应用 [M]. 北京：人民邮电出版社，

2016. (Lang Weimin. Mobile Cloud Computing: Architectures, Algorithms and Applications [M]. Beijing: Posts & Telecom Press, 2016.)

[5] Kamal R. Mobile computing [M]. Oxford: Oxford University Press, 2008.

[6] Meroufel B. , Belalem G. Lightweight coordinated checkpointing in cloud computing [J]. Journal of High Speed Networks, 2014, 20 (3): 131-143

[7] Zhang Wen, Wen Yin, Wu Du. Collaborative task execution in mobile cloud computing under a stochastic wireless channel [J]. IEEE Trans on Wireless Communications, 2015, 14 (1): 81-93.

[8] 黄国盛. 移动 IP 的切换与移动性管理研究 [D]. 长沙: 中南大学, 2010. (Huang Guosheng. Study of handoff and mobility management in mobile IP [D]. ChangSha: Central South University, 2010.)

[9] Zickuhr K. Location-based services [J]. Pew Research, 2013: 1-25.

[10] Al-Molegi A, Jabreel M, Ghaleb B. STF-RNN: space time features-based recurrent neural network for predicting people next location [C]// Computational Intelligence. 2016: 1-7.

[11] Wu Fan, Fu Kun, Wang Yang, *et al.* A Spatial-temporal-semantic neural network algorithm for location prediction on moving objects [J]. Algorithms, 2017, 10 (2): 37.

[12] Zheng Y, Zhang L, Ma Z, *et al.* Recommending friends and locations based on individual location history [J]. ACM Trans on the Web, 2011, 5 (1): 5.

[13] Meetei, K. P. , George, A. handoff management in wireless networks using predictive modelling [C]// Proc of National Conference on Communications. 2011: 1-5.

[14] Feng Zhenni, Zhu Yanmin. A survey on trajectory data mining: techniques and applications [J]. IEEE Access, 2017, 4: 2056-2067.

[15] Liu T, Bahl P, Chlamtac I. Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks [C]. IEEE J. Select. Area Commun, 1998, 16: 922–936.

[16] 薛丹. 用户移动模式的挖掘与位置预测的研究 [D]. 上海: 东华大学, 2005. (Xue Dan. Research on ming the user mobility patterns and location prediction [D]. Shanghai: Donghua University, 2005.)

[17] Jeung H, Liu Q, Shen H T. A hybrid prediction model for moving objects [C]// Proc of the 24th IEEE International Conference on Data Engineering. 2008: 70-79.

[18] Morzy M. Prediction of moving object location based on frequent trajectories [M]. Berlin: Springer Berlin Heidelberg, 2006: 583-592.

[19] Morzy M. Mining frequent trajectories of moving objects for location prediction [M]. Berlin: Springer, 2007: 667-680.